

A	B	C	D	E
1	Term #	The Times's Proposed Search Terms	Total Documents Hit by Term Including Families	OpenAI Objections
2	Unique, De-duplicated Hit Counts for all Terms Combined		546,902	
3				
4	Section A: Terms Relating to the Products' Outputs			
5				
6	1	paywall	861	The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
7	2	antiregurg*	525	The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
8	3	regurg*	86	OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
9	4	(Bing* w/5 index) OR (search w/5 index) OR (search w/5 result*) OR (ranking* w/5 engine) OR ((knowledge OR entity) w/5 graph) OR ((prompt* OR query) w/5 logs) OR (user w/5 (queries OR prompt*))	47,148	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
10	5	Retriev* w/15 ("based model" OR knowledge OR document* OR information)	13,679	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
11	6	(retriev* w/5 augmen*) OR RAG	69	OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
12	7			
13	8	(generat* OR ground* OR synthetic OR semantic OR deriv*) w/25 search*	16,979	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
14	9	REDACTED	5,277	OpenAI provided reasonable search terms that are designed to capture relevant documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
15				
16	Section B: Terms Related to GenAI and Journalism			
17				
18	11	"NY Times" OR NYT OR "New York Times" OR "The Times" OR *nytimes* OR "The Athletic" OR *theathletic* OR *wirecutter*	11,572	OpenAI provided reasonable search terms that are designed to capture relevant documents. This term is overbroad and likely to capture a large volume of irrelevant documents.
19	12	(ethic* w/5 committee) OR (ethic* w/5 audit) OR (ethic* w/5 review) OR "ethic* Team" OR (red w/5 team*) OR "AI impact assessment" OR Aether OR "pre-deployment reviews" OR "stress testing" OR "responsible AI" OR RAI OR "Python Risk Identification Tool" OR PyRIT OR (compliance w/5 audit) OR (compliance w/5 committee)	28,300	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
20	13	(Traffic OR search* OR clickthrough OR click-through OR "click through" OR CTR) w/25 (*GPT* OR dv3 OR dv-3 OR davinci-3 OR genai OR generative OR Bing OR Copilot OR Azure OR genai OR SLM OR Microsoft OR Msft OR Prometheus OR Orchestrat*)	33,819	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
21				

A	B	C	D	E
1	Term #	The Times's Proposed Search Terms	Total Documents Hit by Term Including Families	OpenAI Objections
22	Section C: Terms Relating to The Times's Legal Claims			
23				
24	14	(Concern* OR importan* OR valu* OR need* OR neces* OR *legal* OR *law* OR complain* OR claim* OR infring* OR violat* OR right* OR "intellectual property" OR IP) w/50 (*gpt* OR dv3 OR generative OR genai OR SLM OR Bing OR Copilot OR Azure OR Microsoft OR Msft OR Prometheus OR Orchestrat*)	147,731	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
25	15	Copyright* OR © OR © OR DRM OR "digital rights management" OR (fair w/5 use)	25,747	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
26	16	(Trademark* OR TM OR Lanham OR dilut* OR tarnish* OR ™ OR &trade OR ® OR ®) w/25 (*gpt* OR dv3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Microsoft OR Msft)	1,437	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
27	17	Copy* w/5 Shield	3	The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
28	18	Indemni*	3,226	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
29	19	((IP OR "intellectual property" OR copy*) AND (polic* OR procedure* OR practice* OR conduct* OR action* OR guidance* OR guideline* OR SOP OR "standing operating procedure")) w/25 (*gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Microsoft OR Msft OR transformer OR "neural network*" OR "machine learning" OR Prometheus OR Orchestrat*)	1,045	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
30	20	CMI OR DMCA OR 1202 OR watermark* OR (copyright w/5 management)	19,492	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
31	21	data w/25 filter*	18,890	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
32	22	(misuse w/3 detect*) OR (domain w/3 filter) OR (action w/3 monitor*) OR (us* w/3 polic*)	16,258	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
33	23	(sketchy OR *reliab* OR *trust*) w/15 (data OR content* OR safety OR polic*)	24,378	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
34	24	(content OR safety OR policy) w/25 (filter* OR mask*)	17,085	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
35				
36	Section D: Terms Relating to Model Training			

A	B	C	D	E
1	Term #	The Times's Proposed Search Terms	Total Documents Hit by Term Including Families	OpenAI Objections
37				
38	25	(*train* OR "fine tun*" OR fine-tun* OR finetun* OR instruct* OR *param* OR corpus OR corpora OR token* OR RLHF OR "proximal policy optimization" OR "data set" OR dataset* OR "test data") w/25 (discriminator OR curricul* OR classif* OR *score* OR frequen* OR contamination* OR duplicat* OR regression* OR cluster* OR feature* OR label* OR attribute* OR categ* OR value* OR variable* OR inference OR distillation OR recipe*)	221,351	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why the ability to inspect OpenAI's training data and source code or documents provided through targeted collections are insufficient.
39	26	(*train* OR "fine tun*" OR fine-tun* OR finetun* OR instruct* OR *param* OR corpus OR corpora OR token* OR RLHF OR "proximal policy optimization" OR "data set" OR dataset* OR "test data") w/25 (rule* OR guidance* OR weight* OR perplexity OR ROUGE OR F1 OR classif* OR benchmark OR whitelist OR blacklist OR safeguard* OR guardrail* OR quality OR recitation* OR scrap* OR checkpoint)	130,910	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why the ability to inspect OpenAI's training data and source code or documents provided through targeted collections are insufficient.
40	27	data w/10 (clean* OR *process* OR curat* OR select* OR augment* OR samp* OR gather* OR copy* OR publish*)	71,644	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agree
41	28	(CC NOT w/25 (.com OR (To AND BCC))) w/50 (data OR content OR *train* OR dataset* OR *gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Prometheus OR Orchestrat*)	123,352	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agree
42	29	(train* OR stage* OR layer*) w/5 (learn* OR iteration OR cycle OR batch OR epoch)	30,177	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why the ability to inspect OpenAI's training data and source code or documents provided through targeted collections are insufficient.
43	30	C4	13,205	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why the ability to inspect OpenAI's training data and source code or documents provided through targeted collections are insufficient.
44	31	webtext* OR WT OR WT2	5,792	OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why the ability to inspect OpenAI's training data and source code or documents provided through targeted collections are insufficient.
45	32			
46	33	(*gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai OR Prometheus OR Orchestrat*) w/50 (journali* OR news* OR report* OR info* OR knowledge OR article* OR story OR stories OR publish* OR headline* OR cover* OR syndicat* OR podcast* OR lede OR scoop* OR editorial OR opinon)	78,357	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
47	34	(impact* w/5 assess*) w/25 (*gpt* OR dv3 OR dv-3 OR davinci-3 OR genai OR generative OR SLM OR Bing OR Copilot OR Azure)	44	The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.

A	B	C	D	E
1	Term #	The Times's Proposed Search Terms	Total Documents Hit by Term Including Families	OpenAI Objections
48	35	(*gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR SLM OR Bing OR Copilot OR Azure OR genai) w/5 metric*	2,512	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
49				
50		Section E: Terms Relating to Technical Documentation and Processes		For all documents in Section E, Plaintiff has not explained why the ability to inspect OpenAI's training data and source code or documents provided through targeted collections are insufficient. These topics are not appropriately addressed via search terms. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
51				
52	36	(reward OR policy) w/10 model*	18,701	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
53	37	(few OR one OR zero) w/10 shot*	27,764	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
54	38	(training OR stag* OR layered) w/5 (learn* OR iteration OR cycle OR batch OR epoch OR curve)	21,245	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
55	39	"model completion"	705	See general objection to Section E.
56	40	data w/25 scrap*	5,539	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
57	41	("validation data" OR "The Pile" OR preprocessing OR "test set" OR "training set" OR "evaluation set" OR "text extraction" OR "semantic analysis" OR "source attribution") w/25 (discriminator OR curricul* OR classif* OR *score* OR frequen* OR contamination* OR duplicat* OR regression* OR cluster* OR feature* OR label* OR attribute* OR categ* OR value* OR variable* OR inference OR distillation OR recipe*)	3,756	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
58	42	("validation data" OR "The Pile" OR preprocessing OR "test set" OR "training set" OR "evaluation set" OR "text extraction" OR "semantic analysis" OR "source attribution") w/25 (rule* OR guidance* OR weight* OR perplexity OR ROUGE OR F1 OR classif* OR benchmark OR whitelist OR blacklist OR safeguard* OR guardrail* OR quality OR recitation* OR scrap* OR checkpoint)	1,676	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
59	43	KPI OR PRD OR (performance w/25 indicator)	73	See general objection to Section E.
60	44	KPI OR PRD OR (performance w/25 indicator)	73	See general objection to Section E.
61	45	measur* w/5 pipe*	130	See general objection to Section E.
62	46	(model w/25 card) OR "product brief" OR "product requirement" OR "data card" OR datasheet OR "data doc"	430	See general objection to Section E.
63	47	(product* OR design* OR technical OR architecture OR system* OR launch*) w/10 (specification* OR requirement* OR brief* OR report* OR chart OR log)	42,250	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
64	48	embed* w/15 (function OR space OR vector OR index* OR indices)	9,855	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
65	49	relevance OR "query tuning" OR (source* w/5 (trust* OR credib*)) OR (standard* w/5 editor*) OR (criteri* w/5 (exclusion OR inclusion))	15,904	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
66	50	REDACTED	1,299	See general objection to Section E.
67	51	Azure AND compute*	20,934	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.

A	B	C	D	E
1	Term #	The Times's Proposed Search Terms	Total Documents Hit by Term Including Families	OpenAI Objections
68	52	(journali* OR news* OR report* OR info* OR knowledge OR article* OR story OR stories OR publish* OR headline* OR cover* OR syndicat*) AND ((data* w/5 quality) OR (high w/5 quality))	34,433	See general objection to Section E. This term is overbroad and likely to capture a large volume of irrelevant documents.
69				
70		Section F: Terms Relating to Collaboration with Microsoft		
71				
72	53	REDACTED	13,675	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
73	54	REDACTED	834	OpenAI provided adequate alternative search terms for this request. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
74	55	crescendo	167	OpenAI provided adequate alternative search terms for this request. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
75	56			
76	57	mail(*@microsoft.com) AND (data* OR train* OR *gpt* OR dv3 OR dv-3 OR davinci-3 OR generative OR Genai OR SLM OR Bing OR Copilot OR Azure OR *compute*)	13,645	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
77	58	MS w/25 (data* OR train*)	3,719	This term is overbroad and likely to capture a large volume of irrelevant documents. OpenAI provided reasonable search terms that are designed to capture relevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why the ability to inspect OpenAI's training data and source code or documents provided through targeted collections are insufficient.
78				
79		Section G: Search Terms Related to the Market for and Value of Content		
80				
81	59	(*Licens* OR agreement* OR compensat* OR pay OR paid) AND (*gpt* OR dv3 OR dv-3 OR davinci-3 OR genai OR generative OR SLM OR Bing OR Copilot OR Azure OR Prometheus OR Orchestrat*)	46,947	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
82	60	(license OR agreement) AND (train* w/15 model*)	6,127	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
83	61	(public* OR "open source" OR "creative commons" OR government OR 1920 OR 1921 OR 1922 OR 1923 OR 1924 OR 1925 OR 1926 OR 1927 OR 1928) w/25 (*gpt* OR dv3 OR genai OR generative OR SLM OR Bing OR Copilot OR Azure)	16,071	This term is overbroad and likely to capture a large volume of irrelevant documents. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents.
84	62	(MAI OR MAI-1 OR Phi OR Phi-2 OR Phi-3) w/50 (azure OR CPU OR GPU OR compute* OR invest* OR financ* OR profit* OR revenue* OR stake OR money OR process* OR server* OR hardware)	515	This term is overbroad and likely to capture a large volume of irrelevant documents. Additionally, OpenAI provided adequate alternative search terms for this request. The Times has not identified any RFP to which this term relates and to which OpenAI agreed to produce documents. Plaintiff has not explained why documents provided through targeted collections are insufficient.
85	63			
86				

A	B	C	D	E
1	Term #	The Times's Proposed Search Terms	Total Documents Hit by Term Including Families	OpenAI Objections
87	Section H: Search Terms OpenAI Seemingly Re-Used from Other Cases			The Times requested documents related to OpenAI's training data in RFP Nos. 3, 14, 17, 33, and 36.
88				
89	64	See general objection to Section H.		
90	65	See general objection to Section H.		
91	66	See general objection to Section H.		
92	67	See general objection to Section H.		
93	68	See general objection to Section H.		
94	69	See general objection to Section H.		
95	70	See general objection to Section H.		
96	71	See general objection to Section H.		
97	72	See general objection to Section H.		
98	73	See general objection to Section H.		
99	74	See general objection to Section H.		